Multimodal Deep Learning for Pediatric Mild Traumatic Brain Injury Detection

Badhan Mazumder[1], Deepan Krishna Tripathy[2], Keith Owen Yeates[3], Miriam H. Beauchamp [4], William Craig [5], Quynh Doan [6], Stephen B. Freedman [7], Catherine Lebel [8], Roger Zemek [9], Ashley L. Ware[10], Dong Hye Ye[11]
[1,2,11]*Department of Computer Science, Georgia State University, Atlanta, GA 30302, USA*
[3]*Department of Psychology, University of Calgary, Calgary, Alberta, Canada*
[4]*Department of Psychology, University of Montreal and*
CHU Sainte-Justine Hospital Research Center, Montreal, Quebec, Canada
[5]*University of Alberta and Stollery Children's Hospital, Edmonton, Alberta, Canada*
[6]*Pediatric Emergency Medicine, University of British Columbia, Vancouver, British Columbia, Canada*
[7]*Departments of Pediatrics and Emergency Medicine, Cumming School of Medicine,*
University of Calgary, Calgary, Alberta, Canada
[8]*Department of Radiology, University of Calgary, Calgary, Alberta, Canada*
[9]*Department of Pediatrics and Emergency Medicine, University of Ottawa,*
Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada
[10]*Department of Psychology, Georgia State University, Atlanta, GA 30303, USA*
[10]*Department of Neurology, University of Utah, Salt Lake City, Utah 84103, USA*
[11]*Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State, Georgia Tech, and Emory University*
bmazumder1@gsu.edu[1], dtripathy1@gsu.edu[2], kyeates@ucalgary.ca[3], miriam.beauchamp@umontreal.ca[4], wcraig@ualberta.ca[5], qdoan@bcchr.ca[6], stephen.freedman@albertahealthservices.ca[7], clebel@ucalgary.ca[8], rzemek@cheo.on.ca[9], alware@gsu.edu[10], dongye@gsu.edu[11]

***Abstract*—Despite its prevalence, little is known about the pathophysiology of mild traumatic brain injury (mTBI). This makes it difficult for clinicians to accurately diagnose mTBI and to predict outcomes in affected children, thereby highlighting the urgent need to identify novel and efficacious biomarkers of pediatric mTBI. To address this important knowledge gap, this study introduced a multimodal magnetic resonance imaging (MRI) based deep learning approach toward the classification of mTBI as compared with mild orthopedic injury (OI) by considering both structural MRI (sMRI) and diffusion tensor imaging (DTI). Firstly, convolutional features were extracted by employing a pre-trained DenseNet to capture the morphological features of both modalities. Next, by employing Deep Canonical Correlation Analysis (DCCA), distinct features obtained from the sMRI and DTI data were integrated into a multi-modal embedding. The obtained DCCA fused compact multimodal features were then fed to a random forest (RF) classifier that was used to classify mTBI versus mild OI. Additionally, to visualize the intra-individually heterogeneous brain regions that DenseNet most heavily relied upon for making classification, Gradient-weighted Class Activation Mapping (Grad-CAM) was applied to the DenseNet outcomes for both modalities. According to the experimental outcomes on the clinical dataset, the introduced multimodal deep learning strategy improved the classification accuracy by 8.6% (from 75.8% to 84.4%) and 7.8% (from 76.6% to 84.4%) when compared to the unimodal morphological features, as generated from sMRI and DTI.**

***Index Terms*—Mild Traumatic Brain Injury, Deep Learning, Transfer Learning, Multi-modal, Explainable AI**

I. Introduction

Traumatic brain injury (TBI) is highly prevalent and can have negative effects on millions of children each year [1]. Most injuries (85–90%) are classified as mild in severity (mTBI; i.e., concussion) [2], [3]. Accurate diagnosis of pediatric mTBI is significantly hindered by the lack of an evidence-based standard of care, which is largely due to the fact that there is no objective clinical biomarker for reliably identifying mTBI or predicting recovery trajectories in injured children [4]. Comparing with adult mTBI, pediatric mTBI is difficult to detect on clinical neuroimaging because not only the effects are heterogeneous across patients but also the effects are easily confounded by intra-individual differences in brain morphology and maturational changes that occur during typical brain development in children [4].

Advanced MRI techniques, such as structural (i.e., T1-weighted; sMRI) and diffusion tensor imaging (DTI), have demonstrated promising sensitivity to the effects of pediatric mTBI [1]. However, growing evidence suggests that quantitative metrics derived from single MRI modalities and hypothesis-driven statistical approaches do not reliably detect the neurobiological underpinnings of mTBI in early periods post-injury [1], [4]. In contrast, machine learning (ML) and convolutional neural network (CNN) approaches have the capacity to locate discriminative features in a data-driven manner, and have recently been able to successfully detect mTBI in adults [5]–[7]. However, previous research has mostly been done based on a single imaging modality, and none of those data-driven approaches have been applied to pediatric mTBI. Multimodal deep learning strategies can help to detect group differences in neurological populations with subtle neurobiological abnormalities, and could increase both scientific

and clinical understanding of the complex neurobiological underpinnings of pediatric mTBI [7]. Specifically, investigating latent interconnections between various MRI modalities with multimodal deep learning methods could help to identify neurobiological traits of pediatric mTBI that are both specific and sensitive enough to be used clinically. Ultimately, this understanding could be used to increase diagnosis accuracy of mTBI in children and to assist clinicians with prognostication.

In this study, a novel multimodal deep learning strategy was introduced that incorporates morphological features from both sMRI and DTI data to classify pediatric mTBI relative to mild orthopedic injury (OI). Identifying and categorizing the unique and shared attributes of each modality were examined with the primary goal being to boost detection rates of pediatric mTBI. At first, convolutional features were obtained from a pre-trained DenseNet for both modalities. After that, we extracted the highly co-related information from both modalities by fusing the features using Deep Canonical Correlation Analysis (DCCA) and then input the compact multimodal features into a random forest (RF) classifier for the identification of mTBI. Moreover, to identify and visualize the brain regions that drove the classification for each modality using DenseNet, Gradient-weighted Class Activation Mapping (Grad-CAM) was employed. Our main contribution was to develop a novel multimodal deep learning framework that efficiently fuses the highly correlated features from across modalities to obtain an improved accuracy when classifying pediatric mTBI from mild OI and also to provide insights into the neurobiology underlying pediatric mTBI detection.

II. Methodology

The current approach consists of two key stages: modality-wise feature extraction and co-relation based fusion of those features for classifying into mTBI or OI. Fig. 1 demonstrates the overview of our proposed approach.

A. Dataset

The data used in this work were drawn from a subset of the Advancing Concussion Assessment in Pediatrics (A-CAP) study [8] dataset. Briefly, the A-CAP study is considered the largest longitudinal study of children between 8 and 16.99 years of age who were recruited at pediatric hospital emergency departments within 48 hours of sustaining a mild OI or mTBI. Children returned for three additional follow-up assessments: a post-acute assessment ($\approx$ 10 days post-injury) and two chronic assessments (3 and 6 months). All patients completed a post-acute 3T MRI without sedation, and were assigned at a random manner in order to complete a second chronic MRI scan at either 3 or 6 months after sustaining the injury. For this study, the post-acute sMRI (T1-weighted) and DTI scans from 120 children (83 mTBI; 37 OI) at the Calgary site were used. Fractional anisotropy (FA) was generated from DTI and gray matter (GM) density was derived from sMRI using FSL [9]. All GM density were smoothed using a Gaussian kernel with sigma = 4.

B. Morphological feature extraction

To generate morphological features from sMRI (GM) and DTI (FA) scans, a transfer learning strategy was employed since it works as an effective technique to avoid over-fitting and also demonstrated increased performance in case of small medical imaging dataset [10], [11]. Following that, in our work, DenseNet-121 was utilized as a pre-trained network for both GM density and FA maps.

As the input of layer m, conventional CNN takes the output of previous layer $m-1$,

$$Y^{(m)} = G_m(Y^{(m-1)}) \tag{1}$$

where G_m represents a non-linear transformation for the m-th layer. The gradient can then be propagated from the rear to the front layers directly, which could cause the vanishing gradient in the network. As a solution to this challenging issue, the DenseNet with the dense blocks architecture was proposed by Huang et. al. [12] employing equation 2

$$Y^{(m)} = G_m([Y^{(m-K)} + \cdots + Y^{(m-2)} + Y^{(m-1)}]) \tag{2}$$

The connection of m-th layer is not only limited with the previous $m-1$ layer but also with all K previous layers which results in improving the gradient flow in between the layers.

A 3D-CNN with a $7 \times 7 \times 7$ kernel and $3 \times 3 \times 3$ max pooling will analyze input data prior to the information flowing through the initial dense block. Each one of the four dense block contains distinct number of dense layers (6, 12, 24 and 16). For each of the dense layer, it contains two 3D CNN with $1 \times 1 \times 1$ and $3 \times 3 \times 3$ kernel sizes, respectively. To decrease the dimensions of feature maps and increase processing performance, at first the 3D-CNN is employed and then there are 3 transition layers that includes batch normalization, ReLU and average pooling for reducing dimensions in between the dense blocks. The 1024-dimensional output feature vector P_j and W_j following the final GAP (global average pooling) layer were considered as the morphological features describing the spatial properties in both GM and FA map, respectively, of the j-th patient.

C. DCCA based multimodal feature fusion

Instead of simply concatenating deep morphological features extracted from both MRI modalities, we applied Deep Canonical Correlation Analysis (DCCA) proposed by Andrew and colleagues [13] for feature fusion. Through a series of layered nonlinear transformations, DCCA computes representations across both MRI modalities that allowed us to fuse highly co-related features from both GM density and FA map which leads towards more accurate detection.

Let, $F_1 \in \mathbb{R}^{M \times d_1}$ be the instance matrix for the obtained GM density feature vector P and $F_2 \in \mathbb{R}^{M \times d_2}$ be the instance matrix for the FA map obtained feature vector W. Here, d_1 and d_2 represents dimensions of P and W, respectively and M indicates the number of instances. Through the use of several stacked layers of nonlinear transformation, two

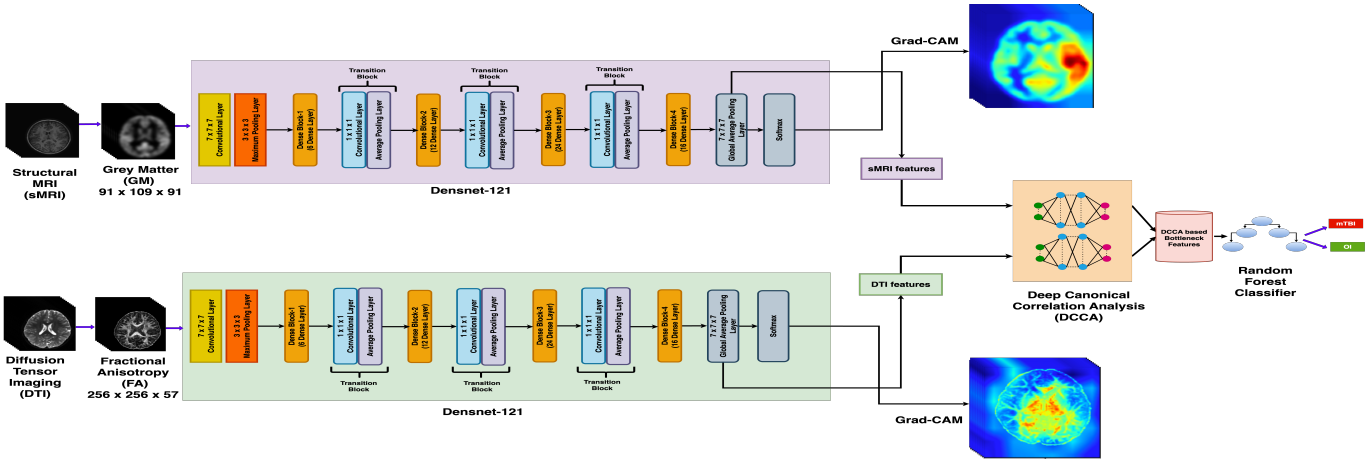

Fig. 1. Overview of the multimodal deep learning approach for pediatric mTBI detection: A pre-trained DenseNet-121 was used to extract 1024 morphological features from the gray matter (GM) density and fractional anisotropy (FA) maps for each patient. Extracted morphological features from both modalities were embedded employing Deep Canonical Correlation Analysis (DCCA) and then input to the random forest classifier for binary classification of mTBI or OI. Grad-CAM was then applied on the obtained classification outcome of DenseNet to visualize and identify the significant brains regions of both MRI modalities responsible for mTBI classification.

neural networks in our DCCA compute representations of both modalities.

Let, $h_1(F_1)$ and $h_2(F_2)$ be the outputs and W_1 and W_2 be the weights of our two neural networks. In order to maximize the CCA objective, W_1 and W_2 were trained by using standard backpropagation:

$$(v_1^*, v_2^*, W_1^*, W_2^*) = \underset{W_1, W_2}{argmax} \quad corr(v_1^T h_1(F_1), v_2^T h_2(F_2)) \tag{3}$$

After the completion of training for both neural networks, the embedded feature C was obtained using a weighted sum fusion method as follows:

$$C = \beta_1 h_1(F_1) + \beta_2 h_2(F_2) \tag{4}$$

where, β_1 and β_2 are weights that satisfies $\beta_1 + \beta_2 = 1$.

D. Classification using Random Forest

For classification purpose the random forest (RF) classifier proposed by Breiman [14] was used as it is parallelizable for efficient computation and comparing with other classifiers, provides high efficiency and also well known for reducing overfitting problem. RF is basically ensemble classification approach that incorporates the bagging technique. As the base classifier RF employs many independent classification and regression trees (CARTs) individually. To classify the input feature vector into mTBI or OI class, each one of the trees provides a unit vote at each input instance for the most popular class and for determining the final label of the class, the majority vote of the trees were taken.

E. Explanation using Grad-CAM

To visualize how the detection were made in our employed deep learning model, we applied Grad-CAM [15] on the outcomes of DenseNet. Before performing global average pooling on the gradients to obtain scalar weights, at first, Grad-CAM determines the gradient of the target decision class corresponding each feature map to avoid the addition of extra layers so that issues with model retraining and performance degradation can be fixed. The class activation map generated from Grad-CAM provided the regions of GM density and FA map that significantly biased DenseNet classification towards mTBI.

III. Results and Discussion

The proposed method was assessed using 5-fold cross-validation, performed with an 80:20 training-testing split ratio on the data from both modalities.

A. Quantitative Evaluation

We measured accuracy, precision and f1-score as quantitative evaluation indices of our binary mTBI classification and enlisted the respective results in Table I. Outcomes in Table I represent that when employing our transfer learning method with the RF classifier in terms of sMRI and DTI individually, accuracy, precision and f1-score ranges between 75.8% to 76.6%, 82.4% to 82.6% and 86% to 86.4%, respectively. The simple concatenation of deep morphological features from these two modalities improved the accuracy, precision and f1-score to 79.8%, 83.6% and 88.6%, respectively which enhanced significantly in terms of all three indices when our introduced DCCA based embedded features were adopted. Employing DCCA based fused features in classification, increased accuracy of 4.6% (79.8% to 84.4%), precision of 0.6% (83.6% to 84.2%) and f1-score of 3% (88.6% to 91.6%) was obtained. This demonstrates the validity of our introduced approach in terms of accounting the two different MRI modalities, since the precise combination of two complementing domains effectively increases the overall performance of mTBI detection. The main advantages of our proposed multimodal technique are two-fold: first, it accounts for fused features from

both modalities which incorporate more information than a single modality; and second, it uses the correlation between sMRI and DTI features early on which better facilitates the mTBI classification.

TABLE I
STATISTICAL ANALYSIS USING 5 FOLD CROSS VALIDATION [UNIT: %] (MEAN±VARIANCE).

Indices	sMRI	DTI	sMRI+DTI (Densenet+RF)	sMRI+DTI (Densenet+DCCA+RF) (proposed)
Accuracy	75.8±0.51	76.6±0.53	79.8±0.03	**84.4±0.05**
Precision	82.4±0.05	82.6±0.02	83.6±0.02	**84.2±0.03**
f1-score	86±0.22	86.4±0.26	88.6±0.02	**91.6±0.01**

B. Qualitative Evaluation

Fig. 2 illustrates representative Grad-CAM generated class activation maps from DenseNet outcomes for both GM density and FA maps for children with mTBI (bottom) and mild OI (top). In GM density, for most of the patients, heterogeneous regions of temporal lobe were found highly responsible for both mild OI and mTBI classification. For FA map, regions surrounding the posterior limb of the internal capsule and posterior thalamic radiation, were found significant for mTBI detection in majority patients. This indicates that minor alterations in these brain regions have high potentiality to affect the deep morphological features extracted from DenseNet which can notably impact the outcome of mTBI detection at an individual patient level.

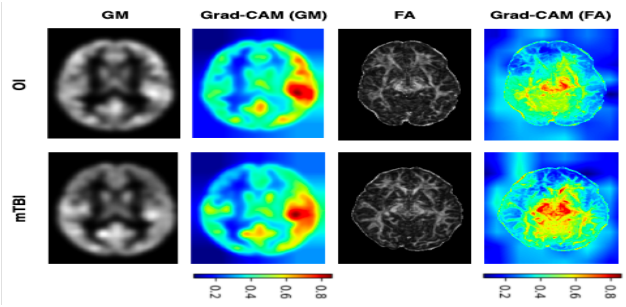

Fig. 2. Activation map using Grad-CAM on the outcome of DenseNet using sMRI (Grey Matter) and DTI (Fractional Anisotropy) individually for mTBI and OI classes. Different regions of temporal lobe seem responsible for sMRI in majority patients. However, in DTI, for most of the patients, regions of posterior limb of the internal capsule and posterior thalamic radiation worked as a significant driving factor for mTBI detection over mild OI.

IV. CONCLUSIONS

This study goes beyond a single data modality concept and considers multimodal MRI features to provide a novel deep learning strategy for the detection of mTBI in children. With increased robustness, extracted deep morphological features from diverse domains were efficiently fused using DCCA. The validity and reliability of our approach is demonstrated by experimental observations on the subset of A-CAP dataset, which demonstrated that, when compared to using a single modality approach, our proposed framework drastically enhances the classification accuracy of pediatric mTBI by incorporating multimodal features based on DCCA. In addition, Grad-CAM was employed to explain the brain regions that were impacted by DenseNet for mTBI detection across both modalities which has the potential for understanding brain biomarkers of pediatric mTBI that could be used to assist with clinical diagnosis.

REFERENCES

[1] A. L. Ware, K. O. Yeates, K. Tang, A. Shukla, A. I. Onicas, S. Guo, N. Goodrich-Hunsaker, N. Abdeen, M. H. Beauchamp, C. Beaulieu *et al.*, "Longitudinal white matter microstructural changes in pediatric mild traumatic brain injury: An a-cap study," *Human Brain Mapping*, vol. 43, no. 12, pp. 3809–3823, 2022.

[2] A. L. Ware, N. J. Goodrich-Hunsaker, C. Lebel, A. Shukla, E. A. Wilde, T. J. Abildskov, E. D. Bigler, D. M. Cohen, L. K. Mihalov, A. Bacevice *et al.*, "Post-acute cortical thickness in children with mild traumatic brain injury versus orthopedic injury," *Journal of neurotrauma*, vol. 37, no. 17, pp. 1892–1901, 2020.

[3] A. L. Ware, A. Shukla, N. J. Goodrich-Hunsaker, C. Lebel, E. A. Wilde, T. J. Abildskov, E. D. Bigler, D. M. Cohen, L. K. Mihalov, A. Bacevice *et al.*, "Post-acute white matter microstructure predicts post-acute and chronic post-concussive symptom severity following mild traumatic brain injury in children," *NeuroImage: Clinical*, vol. 25, p. 102106, 2020.

[4] A. R. Mayer, D. K. Quinn, and C. L. Master, "The spectrum of mild traumatic brain injury: a review," *Neurology*, vol. 89, no. 6, pp. 623–632, 2017.

[5] M. VergaraVictor, R. MayerAndrew, A. KiehlKent *et al.*, "Detection of mild traumatic brain injury by machine learning classification using resting state functional network connectivity and fractional anisotropy," *Journal of neurotrauma*, 2017.

[6] S. Minaee, S. Wang, Y. Wang, S. Chung, X. Wang, E. Fieremans, S. Flanagan, J. Rath, and Y. W. Lui, "Identifying mild traumatic brain injury patients from mr images using bag of visual words," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*. IEEE, 2017, pp. 1–5.

[7] H. Yu, T. Florian, V. Calhoun, and D. H. Ye, "Deep learning from imaging genetics for schizophrenia classification," in *2022 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2022, pp. 3291–3295.

[8] K. O. Yeates, M. Beauchamp, W. Craig, Q. Doan, R. Zemek, B. H. Bjornson, J. Gravel, A. Mikrogianakis, B. Goodyear, N. Abdeen *et al.*, "Advancing concussion assessment in pediatrics (a-cap): a prospective, concurrent cohort, longitudinal study of mild traumatic brain injury in children: study protocol," *BMJ open*, vol. 7, no. 7, p. e017012, 2017.

[9] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney *et al.*, "Advances in functional and structural mr image analysis and implementation as fsl," *Neuroimage*, vol. 23, pp. S208–S219, 2004.

[10] S. H. Gheshlaghi, C. N. E. Kan, and D. H. Ye, "Breast cancer histopathological image classification with adversarial image synthesis," in *2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*. IEEE, 2021, pp. 3387–3390.

[11] N. Lang, D. Saxena, T. Yen, J. Jorns, B. Yu, and D. H. Ye, "Breast cancer magnification-independent multi-class histopathology classification using dual-step model," in *Medical Imaging 2021: Digital Pathology*, vol. 11603. SPIE, 2021, pp. 269–276.

[12] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2016.

[13] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *International conference on machine learning*. PMLR, 2013, pp. 1247–1255.

[14] L. Breiman, "Random forests," *Machine learning*, vol. 45, pp. 5–32, 2001.

[15] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.